PB85-160893

ENHANCED SPECTRUM MATCHING OF MIXED MASS SPECTRA BY MEANS OF QUALITY FACTORS (JOURNAL VERSION)

Environmental Research Laboratory
Athens, GA

1984

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completi

1. REPORT NO. EPA/600/J-84/222	2.
3. PB85-160893	
4. TITLE AND SUBTITLE Enhanced Spectrum Matching of Mixed Mass Spectra by Means of Quality Factors (Journal Version)	5. REPORT DATE 1984 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) W.M. Shackelford and D.M. Cline	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Environmental Research Laboratory U.S. Environmental Protection Agency Athens GA 30613	10. PROGRAM ELEMENT NO. ABHB1C 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency--Athens GA Office of Research and Development Environmental Research Laboratory Athens GA 30613	13. TYPE OF REPORT AND PERIOD COVERED Journal article 14. SPONSORING AGENCY CODE EPA/600/01

15. SUPPLEMENTARY NOTES
Published in Analytica Chimica Acta 164:251-256, October 1984
Contact: W.M. Shackelford, 8-250-3186

16. ABSTRACT

A system of computer programs for recognizing inpure or mixed spectra was assembled that provided automatic subtraction of reference mass spectra from a mixed spectrum. The software was 99% successful in recognizing pure spectra for the cases tested. Subtractions were done on mixed spectra in 70% of the cases tested. Spectrum subtraction enhanced the ability of the matching program to match components of a multicomponent mixture correctly, whereas quality factors were a great aid in the evaluation of the overall match validity.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 7
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

EPA/600/J-84/222
JOURNAL ARTICLE

Analytica Chimica Acta, 164 (1984) 251—256
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

ENHANCED SPECTRUM MATCHING OF MIXED MASS SPECTRA BY MEANS OF QUALITY FACTORS

W. M. SHACKELFORD* and D. M. CLINE

U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, GA 30613 (U.S.A.)

(Received 7th June 1984)

SUMMARY

A system of computer programs for recognizing impure or mixed spectra was assembled that provided automatic subtraction of reference mass spectra from a mixed spectrum. The software was 99% successful in recognizing pure spectra for the cases tested. Subtractions were done on mixed spectra in 70% of the cases tested. Spectrum subtraction enhanced the ability of the matching program to match components of a multicomponent mixture correctly, whereas quality factors were a great aid in the evaluation of the overall match validity.

The matching of mass spectra with a library of reference spectra involves special problems when the spectra acquired from gas chromatography/mass spectrometry (g.c./m.s.) data are not pure. For field samples, impure spectra must be dealt with quite often. Rosenthal [1] has demonstrated that, even when 1-s scans and capillary g.c. columns are used, overlapping components are a virtual certainty for such samples.

Faced with impure spectra as a routine problem, the analyst spends much time trying to unravel both the mixed spectra and the matched results generated from the mixed spectra. In this work, an attempt was made to automate much of this unravelling through the use of quality factors as developed by the McLafferty group [2] and automated spectrum subtraction.

The use of reverse searching to identify components of a mixed spectrum has been reported by several workers [3—5]. The subtraction of a matched spectrum from a mixed spectrum has been reported previously also [6]. An inherent problem in both of these approaches, however, lies in establishing whether the spectrum is pure or impure. Even though match parameters may indicate spectrum purity, the differences in instrument or acquisition conditions between the unknown and reference spectra may result in misleading purity calculations (see Fig. 1).

In this work, quality factors as previously proposed [2] for the determination of purity of reference spectra were applied to the results of spectrum matches by using the probability based matching (PBM) system [7] to estimate spectrum purity. Spectrum subtraction was applied only if the quality

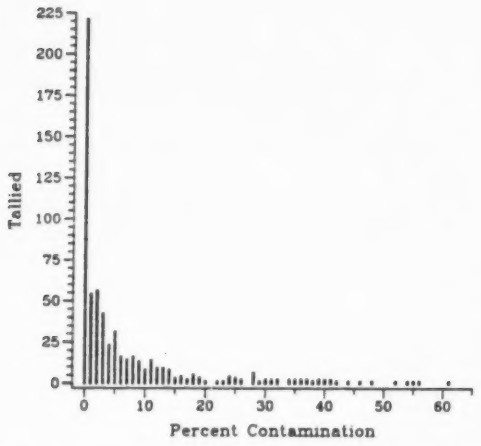

Fig. 1. Distribution of percent contamination for 609 pure spectra calculated by PBM.

factor test was failed. All stages of the matching system were handled automatically.

EXPERIMENTAL

The spectra used for unknowns in this study were obtained from the Research Triangle Institute, Research Triangle Park, NC [8]. All 609 pure spectra represented compounds, the spectra of which were contained in the reference library used for matching in this laboratory [9]. To produce the mixed spectra, a computer program was written to combine linearly and renormalize binary, ternary, and quaternary mixtures. Spectra were mixed in groups of nine. Within each group, spectrum numbers 3 and 6 were mixed to form the binary mixtures. Spectrum numbers 2, 5, and 8 were mixed for the ternary mixtures and spectrum numbers 1, 4, 7, and 9 were mixed for the quaternary mixtures. Thus, from the 609 pure spectra, 67 groups of nine spectra resulted in 603 mixed spectra. Table 1 shows the distribution of mixtures in the set of mixed spectra. As can be seen, component concentrations within mixed spectra ranged from 20% to 80%.

Computer algorithms and hardware

For spectrum matching, the PBM [7] system was used as modified in this laboratory [9]. To calculate the quality factors, code was obtained from the McLafferty group at Cornell University and modified in this laboratory. Only quality factor QF3, which measures contaminant peaks above the molecular weight of the reference compound, and QF4, which measures the extent of illogical neutral losses, were used. Other quality factors discussed earlier [2] do not give as direct a measure of spectrum impurity as QF3 and QF4.

In order to match a spectrum, the PBM was first run and QF3 and QF4

TABLE 1

Distribution of mixed spectra (67 spectra for each case)

Number of components	Ratio	Number of Components	Ratio
2	1:1	3	1:1:1
2	1:2	3	1:2:1
2	1:3	3	1:2:2
2	1:4	4	1:1:1:1
		4	1:1:1:2

were calculated for each of the ten best matches assuming the identity of the unknown spectrum to be each of the compounds generated by the match results. The match results were then scanned to ensure that the best match had more positive than negative match quality. For matches of sufficient quality (see below), the product of QF3 and QF4 was scanned to look for any values less than 0.70. If a value less than 0.70 was found, the best matched reference spectrum (the spectrum with the highest match correlation) was subtracted from the unknown spectrum and the PBM was run on the residual. This process could be repeated for a maximum of four PBM runs.

Because a number of poorly matched compounds can show up in a list of matches for even pure spectra, logic was devised to let the computer judge whether the best spectrum match was of sufficient quality to disregard any subsequent matches that might have the product, QF3XQF4, less than 0.70. The best match for the unknown was the key to further action. If the best match showed K (positive correlation) greater than ΔK (negative correlation) and a calculated contamination less than 30%, the spectrum was deemed pure. If the contamination percentage was greater than 30% but less than 40%, spectrum purity was conceded only if the molecular ion also was found. For cases with contamination percentage greater than 40%, the spectrum was deemed to be contaminated and a match with the quality factor product less than 0.70 was sought. If such a match was found, the best matched spectrum was subtracted from the unknown and the whole process was restarted. If at any point in the process the spectrum was deemed pure, no further processing took place.

Spectrum subtraction was achieved by subtracting the condensed [7] reference spectrum from the unknown spectrum after multiplying the reference spectrum by the purity factor calculated by PBM. Although in earlier work [6] the complete reference spectrum was used for subtraction, storage restrictions made the retention of both a condensed reference data base and the full-spectrum data base on line at one time impractical. Because each condensed spectrum contains between 16 and 26 m/z intensity pairs, it was considered sufficient for this work. After spectrum subtraction occurred, the residual spectrum was renormalized.

All operations were done with a Digital Equipment Corporation PDP 11/70 computer.

RESULTS AND DISCUSSION

To test the use of quality factors as a measure of unknown spectrum purity, both pure and mixed spectra were studied. The collection of 609 pure spectra and 603 mixed spectra were run through the matching routine and statistics were kept to establish the best logic for automatic determination of spectrum purity.

In Fig. 2A, the distribution of values for QF3 are shown for the pure spectra. As expected, essentially all spectra show no contaminant peaks above the molecular weight. In Fig. 2B, the distribution of values for QF4 are shown. Again, although there is some scatter, all values lie essentially near the perfect score of 1.0. Thus, it can be seen that, for pure spectra acquired under routine g.c./m.s. conditions, the quality factors accurately describe purity.

Using quality factors alone, however, does not address the situation in which the components have molecular weights that differ more than the range of illogical neutral losses used in the QF4 calculation, 38 daltons. In these cases, the heavier component will show no interference above its molecular weight and no illogical neutral losses, because the QF4 calculation will not consider losses of greater than 38 daltons. Although the lighter of the components will definitely have QF3 < 1, the heavier component may well look as if it is pure. In the case when the heavier component also is the best matched component, the computer cannot make the decision to subtract without additional data.

To aid the computer at this point, a second parameter is examined to determine spectrum purity; the second parameter is the value calculated by the PBM for contamination percentage. In Fig. 1 the distribution of values for the pure spectra is shown. The 80th percentile of all values for the pure

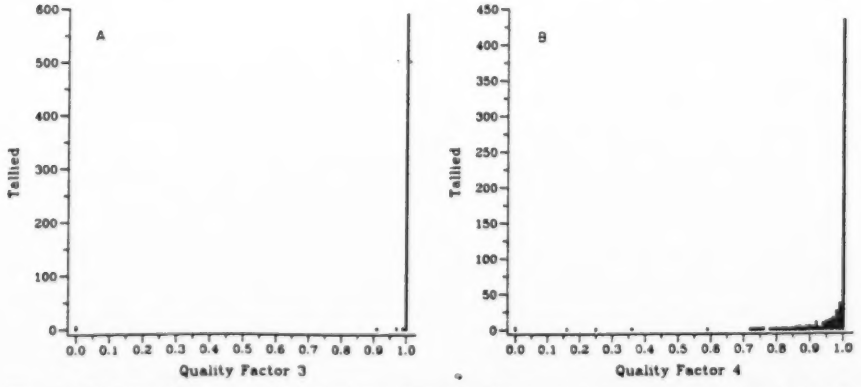

Fig. 2. Distribution of quality factor values for 609 pure spectra: (A) quality factor 3; (B) quality factor 4.

spectra is reached at the 30% level. This was set as the level below which, barring any indications from QF3 and QF4, a spectrum was deemed pure. At intermediate levels, 31—40%, if the molecular ion were present as well, purity was indicated. At contamination levels greater than 40%, all matches were examined for low quality factor values. If any low values were found, the best match in the reference library was subtracted from the unknown spectrum and the process repeated.

Of the 609 pure spectra sent to the matching routine, 605 were seen to be pure by the above logic. Of the 603 mixed spectra sent to the matching routine, 422 were correctly seen to be mixed. The compounds that consistently gave problems to the logic were heavily halogenated aromatics. Because of their relatively high molecular weights and characteristic isotope clusters, these compounds had very low contamination matches and exhibited almost no effects on their quality factors. Overall, although about 30% of the mixed spectra were incorrectly classified as to purity, much computer time was saved by avoiding unnecessary spectrum subtraction of the pure spectra.

In Table 2 are shown the matching results for the pure and mixed spectra. For this work, a correct answer corresponds to a Class 2 [7] match (the correct compound or an isomer differing by a shift of only one carbon position) of the unknown as the best PBM match. Thus, although the reference library contained spectra corresponding to each of the pure spectra, there were cases in which these spectra were not the best match. For pure spectra, 92% were matched.

For the mixtures, the results were promising. As can be seen in Table 2, spectrum subtraction improved the number of compounds that were matched in mixtures over what could be expected for a single matching step. It is evident from Table 2 that, as the number of compounds within the mixture increases or the relative proportion of a compound within a mixture decreases, the ability of the system to retrieve correct matches decreases.

TABLE 2

Spectrum matching results

Number of components	Ratio	Correct answers	Percentage correct	Expected percentage correct
1	N/A	562	92	100
2	1:1	115	86	50
2	1:2	91	68	50
2	1:3	76	57	50
2	1:4	69	51	50
3	1:1:1	109	57	33
3	1:2:1	92	48	33
3	1:2:2	83	43	33
4	1:1:1:1	98	37	25
4	1:1:1:2	89	33	25

This decrease comes from both the inability of the matching program to retrieve a good initial match (in the case of three- and four-component mixtures) as well as the first match appearing pure (in the case of 1:3 and 1:4 binary mixtures).

As shown in Table 2, automatic spectrum subtraction resulted in a 36% increase in identifications for unweighted binary mixtures, a 23% increase for unweighted tertiary mixtures and a 12% increase for unweighted quaternary mixtures over the expected identification rate.

Conclusions

Use of quality factors for automating spectrum subtraction in conjunction with the matching of mixed spectra is a valid approach to save operator time. Improvement in the identification rate of compounds is greatest in binary mixtures, which are by far the most frequently encountered in real samples, but is noticeable in ternary and quaternary mixtures.

We thank Prof. F. W. McLafferty, of Cornell University for the quality factor FORTRAN code, Richard Bowen for programming help and Laura Folwell for collation of results.

REFERENCES

1 D. Rosenthal, Anal. Chem., 54 (1982) 63.
2 D. D. Speck, R. Venkataraghavan and F. W. McLafferty, Org. Mass Spectrom., 13 (1978) 209.
3 F. P. Abramson, Anal. Chem., 47 (1975) 45.
4 F. W. McLafferty, R. H. Hertel and R. D. Villwock, Org. Mass Spectrom., 9 (1974) 690.
5 B. E. Blaisdell and C. C. Sweely, Anal. Chim. Acta, 117 (1980) 18.
6 B. A. Fell, R. Venkataraghavan and F. W. McLafferty, Anal. Chem., 51 (1979) 1945.
7 G. M. Pesyna, R. Venkataraghavan, H. E. Dayringer and F. W. McLafferty, Anal. Chem., 48 (1976) 1362.
8 J. T. Bursey and E. D. Pellizzari, EPA Report EPA-600/S4-83-028, 1983.
9 W. M. Shackelford, D. M. Cline, L. Faas and G. Kurth, Anal. Chim. Acta, 146 (1983) 15.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.